Photosynthesis Research **75:** 277–285, 2003.

Hypothesis

The FMO protein is related to PscA in the reaction center of green sulfur bacteria

John M. Olson[1,*] & Jason Raymond[2]
[1]*Department of Biochemistry and Molecular Biology, Lederle Graduate Research Center, University of Massachusetts, Amherst, MA 01003-4505, USA;* [2]*Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1604, USA;* **Author for correspondence (e-mail: jmo@biochem.umass.edu; fax: +1-413-545-3291)*

Received 29 August 2002; accepted in revised form 23 December 2002

Key words: FMO protein, green sulfur bacteria, PscA, reaction center

Abstract

The Fenna–Matthews–Olson protein is a water-soluble protein found only in green sulfur bacteria. Each subunit contains seven bacteriochlorophyll (BChl) *a* molecules wrapped in a string bag of protein consisting of mostly β sheet. Most other chlorophyll-binding proteins are water-insoluble proteins containing membrane-spanning α helices. We compared an FMO consensus sequence to well-characterized, membrane-bound chlorophyll-binding proteins: L & M (reaction center proteins of proteobacteria), D1 & D2 (reaction center proteins of PS II), CP43 & CP47 (core proteins of PS II), PsaA & PsaB (reaction center proteins of PS I), PscA (reaction center protein of green sulfur bacteria), and PshA (reaction center protein of heliobacteria). We aligned the FMO sequence with the other sequences using the PAM250 matrix modified for *His* binding-site identities and found a signature sequence (LxHHxxxGxFxxF) common to FMO and PscA. (The two *His* residues are BChl *a*. binding sites in FMO.) This signature sequence is part of a 220-residue C-terminal segment with an identity score of 13%. PRSS (Probability of Random Shuffle) analysis showed that the 220-residue alignment is better than 96% of randomized alignments. This evidence supports the hypothesis that FMO protein is related to PscA.

Abbreviations: BChl – bacteriochlorophyll; *Cb.* – *Chlorobium*; FMO – Fenna–Matthews–Olson; *H.* – *Heliobacillus*; *Pr.* – *Prosthecochloris*; *Rb.* – *Rhodobacter*; RC – reaction center; *Syn* – *Synechocystis*

Introduction

The Fenna–Matthews–Olson (FMO) protein is a very unusual, water-soluble chlorophyll protein found only in green sulfur bacteria. It functions as an excitation transfer protein between the chlorosome and the reaction center. Each subunit (365/366 amino acids) contains 7 bacteriochlorophyll (BChl) *a* molecules wrapped in a string bag of protein consisting of 15 strands of β sheet, 6 short lengths of α helix, and a few regions of irregular conformation (Matthews et al. 1979; Li et al. 1997). Most other chlorophyll proteins are water-insoluble proteins containing membrane-spanning α helices. With respect to structure and evolutionary history it is a great mystery where the FMO protein came from After a search of 20 million sequences which yielded no likely homologs, it was decided to compare 11 well-characterized, membrane-bound chlorophyll proteins to the FMO protein to see if we could detect any sequence homologies.

Methods

Sequences were downloaded using GenBank queries from the NCBI website (http://www.ncbi.nih.gov) unless otherwise noted. A consensus FMO protein sequence was constructed from the individual sequences

Table 1. Alignment of FMO consensus with various chlorophyll proteins

Protein	Organism	Residues	Start	End	Identities (score,%)	Similarities (score,%)	J-site matches
L subunit	*Rb. capsulatus*	282	6	366	39 (10)	98 (27)	1
M subunit	*Rb. capsulatus*	307	12	367	33 (9)	100 (27)	0
D1	*Syn* 6803	360	10	389	31 (8)	120 (31)	0
D2	*Syn* 6803	352	1	385	50 (13)	122 (32)	0
CP43	*Syn* 6803	472	2	472	54 (11)	132 (29)	1
CP47	*Syn* 6803	507	43	506	54 (11)	131 (26)	0
PsaA	*Syn* 6803	751	116	751	59 (8)	153 (20)	1
PsaB	*Syn* 6803	731	180	729	62 (8)	150 (21)	1
PshA	*H. mobilis*	609	147	608	49 (8)	147 (24)	0
PscA	*Cb. limicola*	730	127	730	56 (8)	132 (18)	3
PscA	*Cb. tepidum*	731	192	731	60 (8)	148 (20)	3
PscA	Consensus	731	129	731	51 (7)	129 (18)	3

of FMO proteins from *Prosthecochloris (Pr.) aestuarii*, *Chlorobium (Cb.) limicola,* and *Cb. tepidum* (Daurat-Larroque et al. 1986; Hager-Braun et al. 1995; Dracheva et al. 1992). A consensus PscA sequence was constructed from the individual sequences from *Cb. limicola* (Büttner et al. 1992 revised by PEDANT 2001) and *Cb. tepidum* (Eisen et al. 2002). The PAM250 matrix (Dayhoff et al. 1978) was modified by the introduction of the letter J to symbolize BChl *a*-binding *His* residues in FMO. The score for matching J–H pairs was set at + 10, while the score for matching H–H pairs was left at the PAM250 default score of + 6. The gap initiation penalty was set at −10 or −8 and the gap extension penalty at −2. Alignments were generated by PAM250 using a Windows NT-based Pentium III-1GHz computer and Bioedit sequence analysis package (Hall 1999). Global alignments were generated using Clustal W 1.81 (Thompson at al. 1994) using the default Gonnet matrix (Gonnet et al. 1994) with gap initiation penalty of −8 and extention penalty of −2 and *without* the use of the letter J.

Results

Preliminary alignments

The amino acid sequences of L and M subunits from *Rhodobacter (Rb.) capsulatus* (Youvan et al. 1984) were aligned with the consensus sequence for the FMO protein with the gap initiation penalty set at −10. There was one match of a J site (H298) in FMO with H231 in L, but there were no matches with *His* residues in M. The sequences of D1 and D2 subunits from *Synechocystis* 6803 (Kaneko et al. 1995) were also aligned with FMO. There were no matches with *His* residues in either D1 or D2.

When FMO was aligned with CP43 and CP47 from *Synechocystis* 6803 (Chisholm and Williams 1988; Vermaas et al. 1987), there was one match of a J site (H110) in FMO with H132 in CP43, but there were no matches in CP47. When FMO was aligned with PsaA and PsaB from *Synechocystis* 6803 (Smart and McIntosh 1991), there was one match of a J site (H298) in FMO with H629 in PsaA and also a match of H297 in FMO with H595 in PsaB. However, when FMO was aligned with PshA from *Heliobacillus (H.) mobilis* (Liebl et al. 1993; Xiong et al. 1998), there were no J-site matches at all.

When FMO was aligned with PscA from *Cb. limicola* (Büttner et al. 1992 revised by PEDANT 2001), there were three J-site matches (FMO/PscA) at H110/H486, H297/H589, and H298/H590 (data not shown). When FMO was aligned with PscA from *Cb. tepidum* (Eisen et al. 2002), there were three J-site matches at H145/H390, H297/H590, and H298/H591 (data not shown). This alignment is identical to that for *Cb. limicola* from positions 545 to 731(C terminus), but quite different from 129 to 544 (N terminus). When FMO was aligned with the consensus sequence for PscA, the three J-site matches were H110/H487, H297/H590, and H298/H591 as shown in Figure 1. It should be noted that the first J-site match varies in these three alignments. This site is the same in the

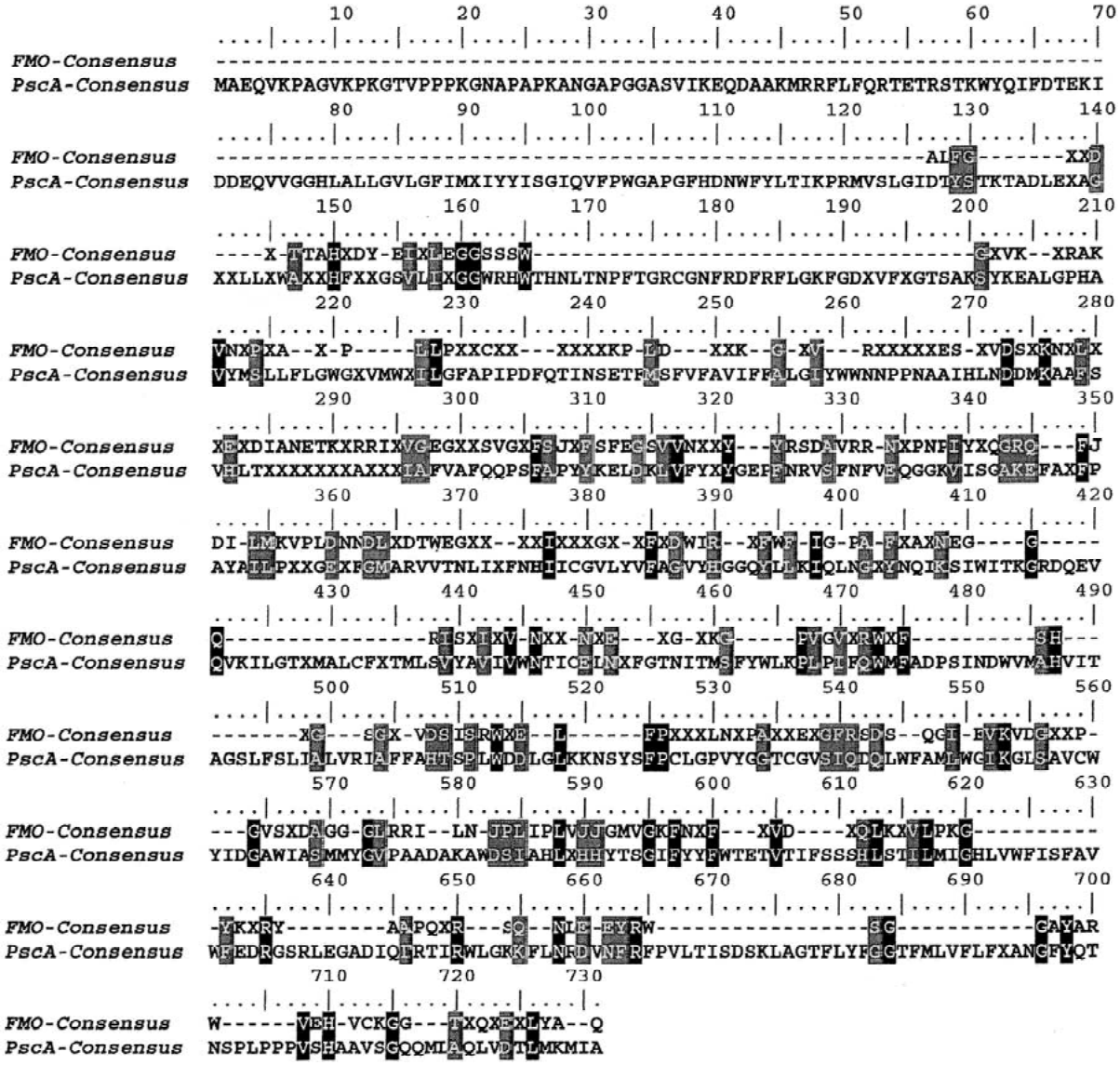

Figure 1. Alignment of the FMO consensus sequence with the PscA consensus sequence generated by PAM250 10/2.

alignments with PscA from *Cb. limicola* and with the consensus sequence, but is different in the alignment with PscA from *Cb. tepidum*.

We concluded from the results summarized in Table 1 that the FMO protein shows identity scores ranging from 7 to 13% and similarity scores from 18 to 32% when compared to 11 chlorophyll proteins from proteobacteria, cyanobacteria, heliobacteria, and green sulfur bacteria. On the basis of the number of identities and similarities, FMO is most similar to PsaA, PsaB, and PscA. (In the alignments of the type I reaction center proteins (PsaA, PsaB, PshA, and PscA) none of the first 120–190 N-terminal residues are aligned with any FMO residue.) On the basis of the number of J-site identities, FMO is most similar to PscA.

Trimmed alignments

In order to reduce the large gaps in the consensus alignment of FMO and PscA (Figure 1) and thereby to increase the identity score to at least 11%, the PscA consensus sequence was systematically shortened from the N-terminus until the alignments met these

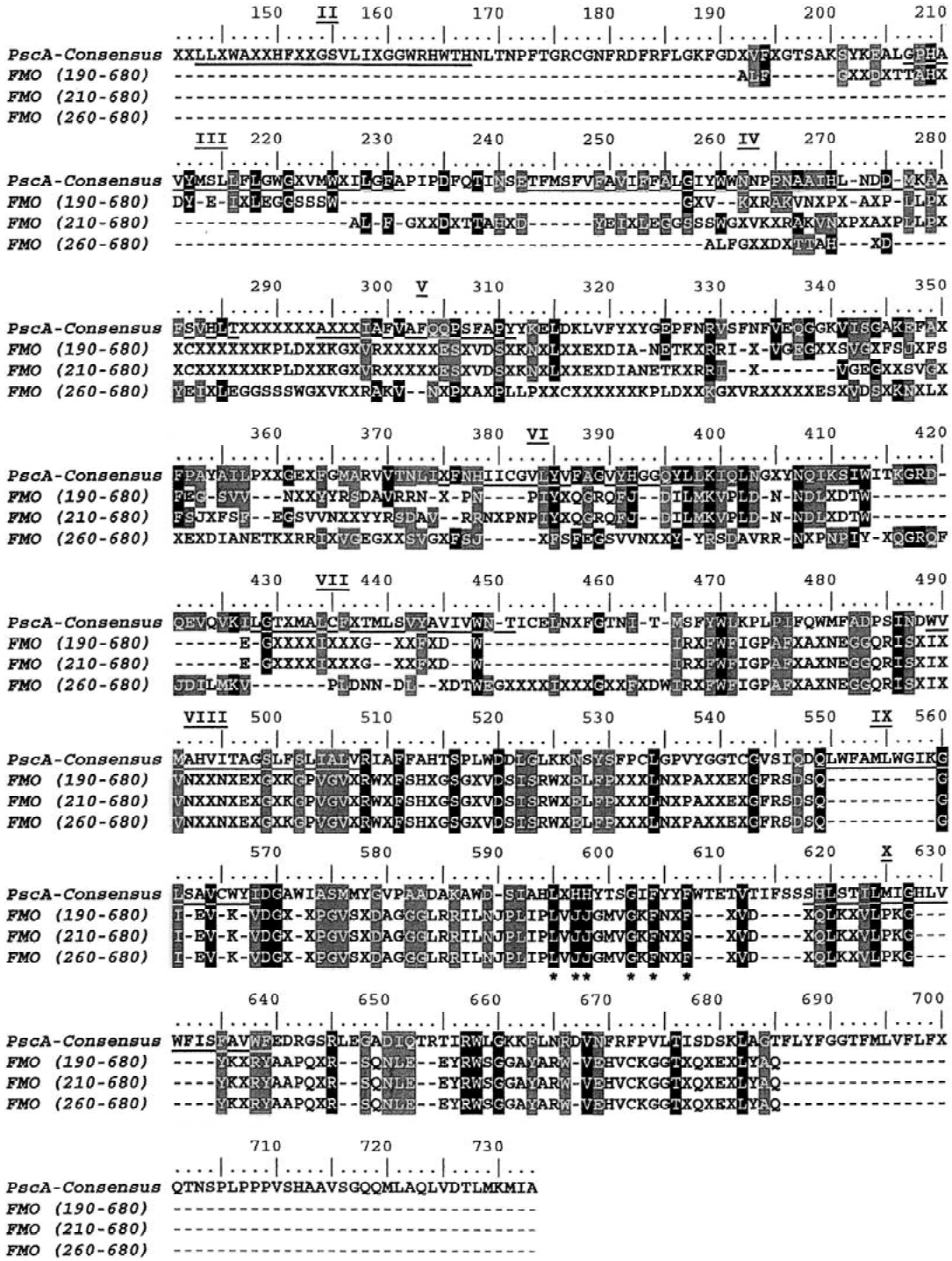

Figure 2. Extended (190–680), intermediate (210–680) and restricted (260–680) maps. Underlined segments in PscA denote α helices labeled II-XI. Identities are shown in black, similarities in gray. Signature sequences are denoted by ***, '–' indicates alignment gaps, 'J' indicates pigment binding histidines where J-site substitutions (see text) were made, and 'X' indicates variable amino acid residues disregarded in constructing the consensus sequence.

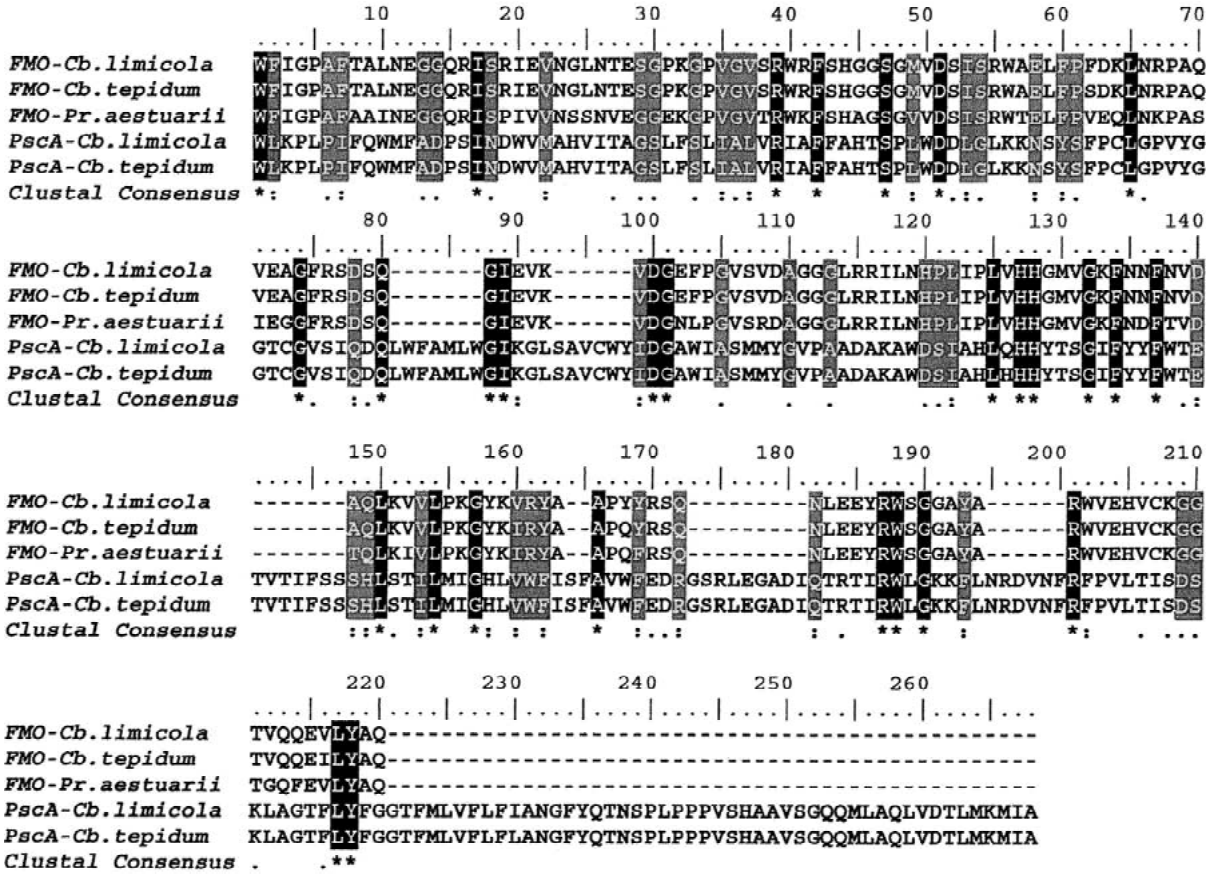

Figure 3. Global alignment of all FMO and PscA sequences (W183/W464 to C termini) using ClustalW 1.82 (PAM250 8/2). Identities are denoted by asterisks (∗), strong similarities by double dots (:), and weak similarities by single dots (.).

criteria. This first took place when the sequence was cut off between T170 and K200. The sequence was then trimmed from the C-terminus (A731) in order to eliminate large gaps and at the same time increase the local identity score at the C-terminus. This was achieved when the sequence was cut off between I670 and L680. The optimum overall alignment was taken to be G190-L680 (*extended alignment* shown in Figure 2) with 53 identities (score (excluding X/X) = 10.8%). This alignment is characterized by a large (31 residue) gap between PscA W225 and G257 (W/W 35 and G/G 67 in Figure 2) with the first binding site at FMO H145/PscA H390 (H/H 200 in Figure 2).

When the N-terminus was further shortened to A210, the large gap disappeared, but the first binding site was not changed. The optimum overall sequence was taken to be A210-L680 (*intermediate alignment* in Figure 2) with 51 identities (score = 10.8%). Further shortening of the N-terminus to W220, F230, and N240 led to failing alignments. However when the N-terminus was cut at A250, another acceptable alignment was obtained. This alignment was further improved by cutting at W260, and the optimum overall sequence was taken to be W260-L680 (*restricted alignment* in Figure 2) with 50 identities (score = 11.7%). In this case, however, the first binding site was shifted to FMO H110/PscA H376 (H/H 117 in Figure 2). Clearly the alignment of the first binding site is not robust, as its position shifts from H110/H376 to H110/H487 to H145/H390 to H145/H490, depending upon the precise alignment of FMO with PscA. The number of identities in the C-terminal region has been substantially increased by trimming the C terminus of PscA.

When J was replaced with H in these alignments, the extended and intermediate alignments remained

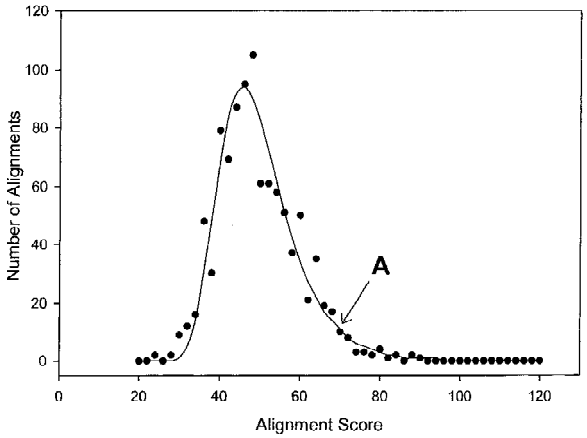

Figure 4. Distribution of scores from 1000 randomized alignments between C-terminal segments of FMO and PscA from *Cb. tepidum*. Observed alignment scores (●) and theoretical distribution (line). The score of the unshuffled alignment (A = 69) is higher than 95.9% of the scores of shuffled alignments.

the same, but the restricted alignment fell apart (data not shown). This shows that the extended and intermediate alignments are more robust than the restricted alignment.

Alignment of all FMO and PscA sequences in the C-terminal region

From comparison of the different trimmed, pairwise alignments between FMO and PscA, we noted that the C-terminal segments of these proteins consistently align with one another so as to preserve the LxHHxxxGxFxxF motif that includes two putative BChl binding sites. These C-terminal segments start at approximately W183 and W464 in FMO and PscA, respectively (*Cb. tepidum* numbering system) and extend to the C termini of both sequences. These regions were used in constructing global alignments of three FMO segments (*Cb. tepidum, Cb. limicola, Pr. aestuarii*) and two PscA segments (*Cb. tepidum, Cb. limicola*) using ClustalW 1.82. The resulting global alignment in Figure 3 agrees in general with the trimmed consensus alignments in Figure 2 and the alignment in Figure 1.

This global alignment was done without modification of the alignment score matrix (no use of J) or the introduction of consensus sequences. The robustly aligning C-terminal region including the LxHHxxxGxFxxF motif appears to represent a good match between the two proteins. To test the robustness of this match, the FMO and PscA sequences from *Cb. tepidum* that were used in the global alignment were shuffled and realigned using the program PRSS (Pearson and Lipman 1988). This program gives an estimate of how well our best alignment would score relative to repetitive randomized alignments of the same sequences. As shown in Figure 4 the score of our optimal alignment is better than 96% of the randomized alignment scores.

We subsequently looked for homologous segments in the N-terminal region of the extended, intermediate, and restricted alignments (Figure 2). We could find no homologous segments common to all three alignments but we did find a 99-residue segment (L315-W413) that is very similar in the extended and intermediate alignments. This region was used to construct a global alignment (Figure 5) of three FMO segments (*Cb. tepidum, Cb. limicola, Pr. aestuarii)* and two PscA segments (*Cb.tepidum, Cb.limicola)*. In this alignment there are 14 identities for a score of 14%. To test the robustness of the 99-residue match, the FMO and PscA sequences from *Cb. tepidum* that were used in the global alignment were shuffled and realigned. The score of our optimal alignment turned out to be no better than 68% of the randomized alignment scores.

To find out whether there is evidence for the LxHHxxxGxFxxF motif in other proteins, we carried out a search against the ERGO database (http://ergo.integratedgenomics.com/ergo/), which consists of 81 fully or partially completed genomes from both eukaryotes and prokaryotes. For LxHHxxxGxFxxF only two protein sequences were found: the PscA sequence from *Cb. tepidum* and a putative gene (631 amino acid open reading frame) from *H. mobilis*. (The ERGO version of *Cb. tepidum* is missing roughly 300kb of genome sequence, which includes the FMO protein, though this motif is clearly present in FMO when searching against *Cb. tepidum* in the TIGR online database http://tigrblast.tigr.org/ufmg/). The open reading from from *H. mobilis* did not have any homologs in the NCBI sequence repository and showed no significant identity to PshA, the heliobacterial reaction center protein. It seems to us both exciting and significant that this pigment binding motif is found among a very small subset of proteins from two organisms with notably similar photosystems. (This motif is also found in the sequence of the delta subunit of a human sodium channel protein BNaC2 (Garcia-Anoveros et al. 1997).)

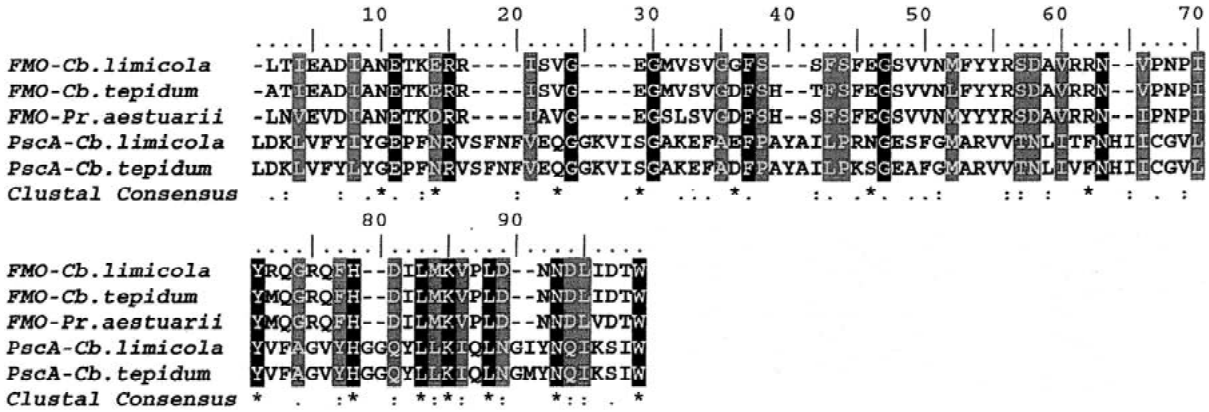

Figure 5. Global alignment of all FMO and PscA sequences (L81/L316–W162/W411) using ClustalW 1.82 (PAM250 8/2). Symbols are the same as in Figure 3.

Discussion

The FMO protein is a water-soluble, mostly β-sheet protein (365/366 amino acids) containing 7 BChl *a* molecules, whereas PscA is a much larger, membrane-bound protein (730/731 amino acids) with 11 putative transmembrane helices, 8 BChl *a* molecules and 2 chlorophyll *a* molecules esterified with Δ 2,6 phytandienol (Kobayashi et al. 2000). In spite of these huge differences in primary and secondary structure, Permentier (2001) has pointed out that the overall optical properties of the RC core complex are remarkably similar to those of the FMO protein, and they have some features in common: (i) the number of BChls per subunit is almost the same (7 *versus* 8); (ii) the pigments are arranged in a seemingly irregular array in both complexes; and (iii) transient absorption spectra of the excited states indicate that on the average the extents of exciton coupling are about the same, suggesting comparable pigment densities in the two complexes.

The utility of sequence alignment in inferring homology is directly undermined by evolutionary divergence and, as in this case, by change in protein function. We presume that the shared motifs observed between FMO and PscA represent conserved sequence elements with similar roles, most likely in pigment binding through *His* ligands. Pairwise alignment of these conserved elements is sensitive to variable parameters such as scoring matrix and gap penalties. Thus it is important to note that once a robust local alignment was arrived at through variation of these pairwise alignment parameters, that this 220-residue local alignment was consistently conserved on multiple alignment using only default parameters based on known patterns of sequence evolution. In the absence of three-dimensional models or intermediately related 'missing link' sequences, these conserved local alignments and shared motifs remain the best thread of evidence for inferring homology. Additionally, the statistical strength of the 220-residue alignment *versus* randomized alignments of the same region is better than would be expected from chance pairwise alignments, as would be expected when aligning non-homologous proteins. On the other hand the 99-residue alignment scores only slightly better than chance *versus* a randomized alignment. Thus we conclude that the primary sequence homology between PscA and FMO in this 99-residue region is not statistically significant.

The closest match between the two proteins is the conserved motif (signature sequence), LxHHxxxGxFxxF. In the FMO protein the sequence LxHH is part of helix 5, while GxxxFxF appears to be part of a connector sequence between helix 5 and β strand 14. One *His* binds BChl 3, and the other binds BChl 7 (Matthews et al. 1979; Li et al. 1997). In PscA the entire signature appears to be included in the N-terminal part of loop 9. We would predict that the two *His* residues might be part of a short α helix in loop 9, and that they might bind two Bchl *a* molecules as in the FMO protein.

We suppose that there might be another BChl *a* binding *His* residue in helix VI of PscA. However, there are two *His* residues in helix VI. In some alignments H376 (PscA) matches up with H110 (β strand

7), while in other alignments H390 (PscA) lines up with H145 (β strand 8). We have no reason to prefer one of these alignments over the other.

The map of FMO on PscA apparently covers at least part of helix IV through loop 10 with many gaps, but it might include helix III and loop 3 as well. It is even possible, but unlikely, that it includes helix II and loop 2 in addition. However, there is no similarity between N terminus + helix I + most of loop 1 and FMO in any alignment.

If the FMO protein is indeed related to PscA as we have shown, there is a strong inference that they share a common ancestor. What kind of common ancestor? From an evolutionary point of view, one can imagine that a photochemically *inactive* chlorophyll protein might have evolved from a photochemically *active* protein, but not the other way around. Therefore, we propose that the common ancestor of FMO and PscA was a primitive RC protein that was membrane bound like PscA. This ancestral RC protein probably consisted of several transmembrane α helices as does PscA. We suppose that the FMO protein evolved from a duplicated ancestral RC protein in several small steps in which the α helical structure of the ancestral protein changed gradually into the β sheet structure of the FMO protein. We assume a gene duplication in which one gene product became photochemically inactive, but was conserved because it transferred excitation energy to the reaction center. Subsequent mutations that destabilized one or both of the terminal α helices would have created a hybrid protein with fewer α helices and more of the peptide exposed to the cytoplasm. We suppose that the exposed peptide reorganized itself into a β sheet structure that retained the ability to transfer excitation energy to the reaction center. After several such mutations the α helices would have disappeared, leaving only the hydrophilic β sheet structure characteristic of the FMO protein.

Conversion of α helical structures to β sheet structures is more common than one might realize. Among the many examples is the conversion of, apomyoglobin from a globular α-helical protein to stable β-sheet fibrils upon heating to 65 °C at pH 9 (Fändrich et al. 2001). Another example is the prion protein that can exist in two forms, a harmless α-helical form (PrP^{c}) and a virulent form (PrP^{sc}) with one or more of the α helices converted into β sheets (Prusiner 1996).

Acknowledgements

The work at Arizona State University was supported by the ASU NASA Astrobiology Institute and by the Exobiology Program of NASA. We thank Prof. John G. Ormerod (University of Oslo) for telling us about the web site, PEDANT, and we thank Prof. Donald A. Bryant (Pennsylvania State University) for sharing the amino acid sequence of PscA from *Cb. tepidum* before publication. We also thank an anonymous reviewer for pointing out that the LxHHxxxGxFxxxF motif is also found in the BNaC2 protein of humans. We especially thank Prof. Robert E. Blankenship (Arizona State University) for continued interest and support throughout this investigation.

References

Büttner M, Xie D-L, Nelson H, Pinther W, Hauska G and Nelson N (1992) Photosynthetic reaction center genes in green sulfur bacteria and in Photosystem 1 are related. Proc Natl Acad Sci USA 89: 8135–8139

Chisholm D and Williams JGK (1988) Nucleotide sequence of psbC, the gene encoding the CP-43 chlorophyll *a*-binding protein of Photosystem II, in the cyanobacterium *Synechocystis* 6803. Plant Mol Biol 10: 293–301

Daurat-Larroque ST, Brew K and Fenna RE (1986) The complete amino acid sequence of a bacteriochlorophyll *a*-protein from *Prosthecochloris aestuarii*. J Biol Chem 261: 3607–3615

Dayhoff MO, Schwartz R and Orcutt BC (1978) Atlas of Protein Sequence and Structure, Vol 5, Suppl 3. National Biomedical Research Foundation, Washington, DC

Dracheva S, Williams JC and Blankenship RE (1992) Sequencing of the FMO-protein from *Chlorobium tepidum*. In: Murata N (ed) Research in Photosynthesis, Vol I, pp 53-56. Kluwer Academic Publishers, Dordrecht, The Netherlands

Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Feldblyum TV, Hansen CL, Craven MB, Radune D, Khouri H, Fujii CY, White O, Venter JC, Volfovsky N, Gruber TM, Ketchum KA, Tettelin H, Bryant DA, and Fraser CM (2002) The complete genome sequence of the green sulfur bacterium *Chlorobium tepidum*. Proc Natl Acad Sci USA 99: 9509–9514

Fändrich M, Fletcher MA and Dobson CM (2001) Amyloid fibrils from muscle myoglobin. Nature 410: 165–166

Garcia-Anoveros J, Derfler B, Neville-Golden J, Hyman BT, Corey DP (1997) BNaC1 and BNaC2 constitute a new family of human neuronal sodium channels related to degenerins and epithelial sodium channels. Proc Natl Acad Sci USA 94: 1459–1464

Gonnet GH, Cohen MA and Benner SA (1994) Analysis of amino-acid substitution during divergent evolution – the 400 by 400 dipeptide substitution matrix. Biochem Biophys Res Commun 199: 489–496

Hager-Braun C, Xie D-L, Jarosch U, Herold E, Büttner M, Zimmermann R, Deutzmann R, Hauska G and Nelson N (1995) Stable photobleaching of P840 in *Chlorobium* reaction center prepara-

tions: presence of the 42-kDa bacteriochlorophyll *a* protein and a 17-kDa polypeptide. Biochemistry 34: 9617–9624

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98

Kanecko T, Tanaka A, Sato S, Kotani H, Sazuka T, Miyajima N, Sugiura M, and Tabata S (1995) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. DNA Res 2: 153–166

Kobayashi M, Oh-oka H, Akutsu S, Akiyama M, Tominaga K, Kise H, Nishida F, Watanabe T, Amesz J, Koizumi M, Ishida N and Kano H (2000) The primary electron acceptor of green sulfur bacteria, bacteriochlorophyll 663, is chlorophyll *a* esterified with Δ2,6-phytadienol. Photosyth Res 63: 269–280

Li Y-F, Zhou W, Blankenship RE and Allen JP (1997) Crystal structure of the bacteriochlorophyll *a* protein from *Chlorobium tepidum*. J Mol Biol 272: 1–16

Liebl U, Mockensturm-Wilson M, Trost JT, Brune DC, Blankenship RE and Vermaas W (1993) Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillis mobilis*: structural implications and relations to other photosystems. Proc Natl Acad Sci USA 90: 7124–7128

Matthews BW, Fenna RE, Bolognesi MC, Schmid MF and Olson JM (1979) Structure of a bacteriochlorophyll *a*-protein from the green photosynthetic bacterium *Prosthecochloris aestuarii*. J Mol Biol 131: 259–285

Pearson WR and Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448

PEDANT (2001) http://pedant.mips.biochem.mpg.de/index.html

Permentier HP (2001) Light-harvesting and core complexes of anoxygenic phototrophic bacteria. PhD dissertation, University of Leiden, The Netherlands, p 87

Prusiner SB (1996) Molecular biology and pathogenesis of prion diseases. TIBS 21: 482–487

Smart LB and McIntosh L (1991) Expression of photosynthesis genes in the cyanobacterium *Synechocystis* sp. PCC6803: psaA-psaB and psbA transcripts accumulate in dark-grown gells. Plant Mol Biol 17: 959–971

Thompson JD, Higgins DG and Gibson TJ (1994) ClustalW – improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Vermaas WFJ, Williams JGK and Arntzen CJ (1987) Sequencing and modification of *psbB*, the gene encoding the CP-47 protein of Photosystem II, in the cyanobacterium *Synechocystis* 6803. Plant Mol Biol 8: 317–326

Xiong J, Inoue K and Bauer CE (1998) Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis.* Proc Natl Acad Sci USA 95: 14851–14856

Youvan DC, Bylina EJ, Alberti M, Begusch H and Hearst JE (1984) Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from *R. capsulata*. Cell 37: 949–957